NPS-OR-97-013

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Realibility Estimation Based Upon
Test Plan Results

by

Robert R. Read

August 1997

Approved for public release; distribution is unlimited.

Prepared for: NAVSURFWARCENDIV Crane
Crane, IN 47522-5001

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943-5000

This report was prepared for and funded by NAVSURFWARCENDIV Crane, Crane, IN 47522-5001.

This report was prepared by:

ROBERT R. READ
Professor of Operations Research

Reviewed by:

RICHARD E. ROSENTHAL
Chairman
Department of Operations Research

GERALD G. BROWN
Associate Chairman for Research
Department of Operations Research

Released by:

DAVID W. NETZER
Associate Provost and Dean of Research

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
August 1997

3. REPORT TYPE AND DATES COVERED
Technical

4. TITLE AND SUBTITLE
Reliability Estimation Based Upon Test Plan Results

5. FUNDING NUMBERS
N0016497WR42204

6. AUTHOR(S)
Robert R. Read

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-OR-97-013

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander, NAVSURFWARCENDIV Crane
(Code 0563, Bldg 2701) 300 Highway 361
Crane, IN 47522-5001

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The report contains a brief summary of aspects of the Maximus reliability point and interval estimation technique as it has been applied to the reliability of a device whose surveillance tests contain a succession of stressful environments. The lower confidence limits appear to be ultra conservative. It is argued that the serial system approach to modeling of the cumulative effect of the stresses is inappropriate. Some alternative modeling is suggested. Some numerical comparisons are offered.

14. SUBJECT TERMS
Reliability, Ammunition

15. NUMBER OF PAGES
29

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

RELIABILITY ESTIMATION BASED UPON TEST PLAN RESULTS

Robert R. Read

Operations Research Deptartment
Naval Postgraduate School
Monterey, CA 93943 USA

Abstract

The report contains a brief summary of aspects of the Maximus reliability point and interval estimation technique as it has been applied to the reliability of a device whose surveillance tests contain a succession of stressful environments. The lower confidence limits appear to be ultra conservative. It is argued that the serial system approach to modeling of the cumulative effect of the stresses is inappropriate. Some alternative modeling is suggested. Some numerical comparisons are offered.

1. Introduction.

The Naval Surface Warfare Center in Crane, Indiana (NSWC - Crane) is engaged in a number of reliability assessment programs for ammunition and pyrotechnic devices. Often reliability figures must be developed from information derived from diverse sources. It is common to "bench" test the components and subsystems individually in order to understand the sources of unreliability. Such tests can be of the binary type, success or failure, and the component reliability information is integrated into a computation of the system reliability.

Computational methodologies have been developed to support these activities [2,3,6, 7,9,14]. Certain aspects are of immediate interest. Complicated systems are decomposed.

Those collections of components that utilize their units in either a purely series or purely parallel way are identified as subsystems. Their reliabilities are estimated from the component test data by either forming products of success rates (series case), or complementing the products of failure rates (parallel case). These subsystems are then viewed as pseudo components whose performances are treated as Bernoulli random variables with success rate estimated as just stated. Since a number of component tests were performed to produce these rates, not all having the same sample numbers, rules [8,9,14] have been developed to convert the collection of component sample numbers into a single number which is used to represent the number of times that the pseudo subsystem has been tested. Further, this number is then multiplied by the estimated reliability in order to obtain an 'equivalent number of successful tests'. Of course this latter number need not be an integer, but still it is regarded as being a binomial random variable in a sense to be described.

The technique cascades. Several such pseudo components combined in either purely series or purely parallel ways are given like treatment; sample sizes and success counts of these new pseudo components are computed as before. Neither of these two quantities need be integral. Nonetheless, it is convenient to speak of them as sample sizes and binomial random variables. Only the ratio is required to produce a success rate, and the Incomplete Beta function can be used to compute confidence intervals. Thus, neither the sample 'size' nor the 'number' of success need be integers. Lower confidence bounds (lcb) of the ultimate system reliability are of particular interest. Generally the lcb's are conservative.

Ideas similar to those just described are also being applied to treat the 'lot acceptance' and 'shelf life' problems; that is, the problem of the initial acceptance of a freshly produced lot and, at later times, the (periodic) surveillance testing of the accepted lots to determine if time, treatment and the environment have affected performance. Lots that

may have deteriorated to below acceptable levels need to be identified and discarded. To this end the test plans contain provision for subjecting the items to sets of specified treatments designed to provide stress and hasten the exposure of defects. The test plans specify the partitioning of the selected sample into subsets such that the items in a single subset receive one and only one of the stresses. No stress at all is usually one of the subsets. Testing ultimately results in the destruction of the items.

Although no single item is subjected to more than one of the stresses, there is a requirement that the reliability estimates refer to items that have received all of the stresses. This requirement is approached with the same methodology indicated in paragraphs two and three above, i.e. treat the various stresses as components of the system. In this case they are treated as independent components of a serial system. The lcb's produced when so doing appear to be overly conservative.

The present report reviews this situation as it pertains to the marine smoke and illumination signal, MK 124 MOD 0, and presents some alternatives that may yield more realistic estimates. Section 2 contains a description of the performance characteristics of the device and the nature of the test plan. The mathematical modeling and notation are presented in Section 3. Section 4 contains summaries of six methods for treating the reliability calculation; two from the Sandia approach [2,14], two from the L.J. Gleser [6] approach, and two by the author. Next, the six methods are subjected to numerical comparison using two test cases in Section 5. This section also contains some simple modeling that could lead to improvement of the lcb's, but at the cost of heavier computation. Such computations have not been pursued. Finally some log linear modeling is proposed in Section 6. This is done to show how one could exploit more of the information that is contained in the test results and to suggest ways to set decision rules in anticipation of changes in the testing requirements. It is expected that the changes will not permit any failures at all in order to accept a lot. A summary section follows and contains

commentary on the appropriateness of the methods. Several appendices are attached that document a variety of details that appear in support of the report.

2. Description of Item Characteristics and Test Plan.

The device is approximately 5.408 inches long and 1.7 inches in diameter. It weighs about half a pound. A flare candle is at one end and a smoke candle is at the other. There are five performance characteristics to be met by the signal:

a.	Display color:	Exhibit orange smoke from one end and a red flare from the other.
b.	Function:	Ignite and produce the displays.
c.	Delay:	No more than three seconds from initiation to the generation of display.
d.	Display Times:	The display time begins after the delay time ends; minimal and maximal display times are specified for each of the two types of signal.
e.	Safety:	The igniter will not separate from the container during function (b).

These are grouped and an item test can have one of four outcomes. That is, a test failure will be marked as being in exactly one of the first three outcomes; the fourth is for a successful test. These outcomes are:

a&e.	Color display or safety failure.
b.	Ignition failure.
c&d.	Delay or burning time specification failure.
suc.	Success; no functioning out of zone.

Prior to testing, an item is subjected to one of the following treatments whose details are specified in [12]. It must withstand them without exploding or burning.

i. Five foot drop.
ii. Forty foot drop.
iii. Vibration.
iv. Temperature and Humidity.
v. High Temperature.
vi. Low temperature.
vii. No stress.

The system test procedures include visual inspection and X-ray prior to proceeding with the ultimate performance tests which are destructive in nature. Of course the inspections are non destructive and when they identify defects in the samples the lots are rejected out of hand. No further testing is done. No reliability computations are required. For those lots that pass the non destructive phase, there is the requirement that reliability figures should reflect the item's ability to perform after having been subjected to all of the stressful situations prescribed.

A random sample of N items is selected from the lot. They are partitioned into k groups of size n_i for $i = 1,\ldots, k$. Thus each sample unit is subjected to exactly one of the treatments. Further, the test plan specifies the threshold for the allowable number of failures in each treatment group and by failure type. The common way is to provide a pair of values separated by a 'slash', e.g. 1/2, which means that a successful test in an outcome type is allowed to suffer zero or one failures among its n_i samples, but must be designated as an item failure if it fails in two or more cases. An excerpt of a test plan sheet is exhibited in Appendix A. For this plan $k = 7$. The present description has been curtailed to show only those aspects that may be of immediate use in the reliability computations. Full details can be found in [12].

3. The Reliability Model.

The goal is to produce point estimates and lower confidence bounds for the reliability of the device, in this case the marine signal. The definition of reliability used is that the item must perform acceptably having been subjected to all of the specified stresses.

The reliability computations presently in use [6,14] reflect models that assume a series system with independent components. Let us examine two of these models in order to understand what they reflect. Much notation and jargon will be borrowed from Gleser [6].

Suppose there are $k-1$ stress treatments and an additional 'no-stress' condition bringing the total to k. This latter situation is tested directly. Gleser would call this the hazard of being produced, and it is marked first in the notation. That is, E_1 is the event that a freshly manufactured device performs acceptably. Let E_i be the event that the device survives the i^{th} stress environment for $i = 2,\ldots, k$. The reliability, R, as defined above is the probability of the intersection of the events $E_1,\ldots, E_k$. When a single stress (say the i^{th}) is a treatment prior to a device test, the success probability is the probability of E_1 and E_i. Thus the reliability goal can be represented as

$$\begin{aligned} R &= \Pr\{E_1 \cap E_2 \cap \ldots \cap E_k\} \\ &= \Pr\{E_1\}\Pr\{E_2 \cap E_3 \cap \ldots \cap E_k\}. \end{aligned} \tag{1}$$

When one assumes the independence of the survivability of stresses, then one can express the reliability in terms of estimable quantities

$$\begin{aligned} R &= \Pr\{E_1\}\prod_2^k P\{E_i|E_1\} \\ &= \Pr\{E_1 \cap E_2\}\ldots P\{E_1 \cap E_k\}\Big/\left[\Pr\{E_1\}\right]^{k-2}. \end{aligned} \tag{2}$$

The quantities $R_1 = \Pr\{E_1\}$ and $R_i = \Pr\{E_1 \text{ and } E_i\}$ are directly estimable from the test using the proportion of successes. The interpretation depends upon what is meant by 'surviving an environment'. In the large, it means acceptable performance after having been treated with the stress. If the conditional probabilities are not unity then some deterioration has taken place. We know not how to measure the deterioration nor how much of it is acceptable. The assumption of independence is made as an expedient since no modeling of this deterioration has been proposed. There is a latent "coherence" assumption that says that application of the stresses do not improve the reliability of the device.

4. Methods.

Three pairs of methods are outlined in this section; Sandia, Gleser, and NPS. The last is proposed by the author. The first two have been proposed in earlier times. It seems useful to summarize them all at once and in a common notation. Also, there are two versions of each (denoted I and II).

Sandia: Let q_1, q_2, ..., q_k be the failure probabilities of the devices after subjection to the various stresses. That is,

$$q_i = \Pr\left\{E_i^c\right\} \quad \text{for } i = 1,2,\ldots,k.$$

Then

$$\begin{aligned} R_i &= \Pr\{E_1 \cap E_i\} = 1 - \Pr\left\{E_1^c \cup E_i^c\right\} \\ &= 1 - q_1 - q_i + \Pr\left\{E_1^c \cap E_i^c\right\}. \end{aligned} \tag{3}$$

Assume the last term is negligible

$$\Pr\left\{E_1^c \cap E_i^c\right\} \approx 0, \text{ for all } i = 2, \ldots, k. \tag{4}$$

The independence assumption allows us to write

$$\begin{aligned} R_2 R_3 \ldots R_k &= \Pr\{E_1 \cap E_2\} \Pr\{E_1 \cap E_3\} \ldots \Pr\{E_1 \cap E_k\} \\ &\approx (1 - q_1 - q_2)(1 - q_1 - q_3)\ldots(1 - q_1 - q_k) \\ &\approx 1 - (k-1)q_1 - \sum_2^k q_i. \end{aligned} \tag{5}$$

On the other hand

$$\begin{aligned} R &= \Pr\{E_1 \cap E_2 \cap \ldots \cap E_k\} \\ &= 1 - \Pr\left\{E_1^c \cup E_2^c \cup \ldots \cup E_k^c\right\} \\ &\approx 1 - \sum_1^k \Pr\left\{E_i^c\right\} + \sum_i \sum_j \Pr\left\{E_i^c \cap E_j^c\right\} \\ &\approx 1 - \sum_1^k q_i. \end{aligned}$$

Combine with (5) and get

$$R \approx \prod_{2}^{k} R_i + (k-2)q_1 . \tag{6}$$

Suppose n_i is the number of items subjected to the i^{th} stress ($i = 1, \ldots, k$) and X_i is the corresponding failure count. We use the estimators $\hat{R}_i = 1 - X_i/n_j$. The approximation (6) suggests the estimator

$$\hat{R} = \hat{R}_p + (k-2)\hat{q}_1 \tag{7}$$

where $\hat{R}_p$ is an estimator for the product $\prod_{2}^{k} R_i$.

It remains to get the lower confidence bound; two methods have been put forth: original [14] and the Gleser [6] modification. We will call them I and II. Both are based upon the use of an equivalent sample size computed from

$$n^* = \hat{R}\left(1-\hat{R}\right) \Big/ \hat{\text{var}}\left(\hat{R}\right) \tag{8}$$

and an equivalent number of successes

$$x^* = n^* \hat{R} . \tag{9}$$

Then the $(1 - \alpha)$ level lower confidence bound is found by solving the equation

$$\alpha = \text{IB}\left(\text{lcb}, x^*, 1 + n^* - x^*\right) \tag{10}$$

where IB is the incomplete Beta function; see Appendix B. The difference in the two methods lies in the manner of estimating the variance of $\hat{R}$ in the denominator of (8).

I. Use the independence of the two terms of (7) in the approximation

$$\hat{\text{var}}\left(\hat{R}\right) \approx \frac{\left[\prod_{2}^{k} \hat{R}_i\right]\left[1 - \prod_{2}^{k} \hat{R}_i\right]}{\min\left(n_i\right)} + (k-2)^2 \frac{\hat{q}_1\left(1-\hat{q}_1\right)}{n_1} . \tag{11}$$

The use of the $\min(n_i)$ comes from the Lindstrom-Madden method [8,14].

II. Gleser has suggested the use of an approximation formula for the variance of the product of independent proportions, see Appendix C. The variance of $\hat{R}_p$, (the first term in (11)) is approximated by eq. (C.1). The estimated form is

$$\operatorname{var}\left(\prod_{2}^{k}\hat{R}_i\right)=\left[\prod_{2}^{k}\hat{R}_i\right]\sum_{j=2}^{k}\left(1-\hat{R}_j\right)\prod_{\substack{2\\ i\neq j}}^{k}\hat{R}_i/n_j. \tag{12}$$

These two methods are called Sandia I and Sandia II.

The Gleser Approach.

Gleser [6] suggested more exploitation of the independence assumption that is implicit in the Sandia analysis. Specifically use estimators in equation (2), i.e.

$$\tilde{R}=\hat{R}_2\hat{R}_3\ldots\hat{R}_k/\left(1-\hat{q}_1\right)^{k-2}. \tag{13}$$

The lower confidence bound uses the Sandia method, but the denominator of (8) is approximated using the delta method for the ratio (13) of two independent random variables. This is outlined in Appendix D and appears as formula (D.3). The approximate variance of (13) is

I. $$\operatorname{Var}\left(\tilde{R}\right)=\frac{\operatorname{Var}\left(\hat{R}_p\right)}{\left(1-q_1\right)^{2k-4}}+(k-2)^2R_p^2\frac{\operatorname{Var}\left(1-q_1\right)}{\left(1-q_1\right)^{2k-2}} \tag{14}$$

and the point estimators $\hat{R}_i$ and $\hat{q}_1$ can be inserted for computation. Then apply (8, 9, 10). This is Gleser I.

Gleser further draws attention to the fact that $\tilde{R}$ is the maximum likelihood estimator for R under the model that $X_i \sim \operatorname{Bin}(n_i, p_i)$ for $i = 1,\ldots, k$ and independent.

For Gleser II the following crude ad hoc method is presented to improve the lower confidence bound. The point estimate is the same, (13). For an approximate lower confidence bound we exploit the fact that the numerator and denominator of (13) are independent random variables. These factors are direct point estimates of R_p and

$(1-q_1)^{k-2}$, respectively, and the ratio of these latter two parameters represent R. We can easily develop an upper confidence bound for $(1-q_1)$ and a lower confidence bound for R_p. The ratio of these bounds, properly adjusted, can provide a lower confidence bound for R. The implementation proceeds as follows. Let

$$\left\{\mathrm{lcb}\left(\hat{R}_p\right) \le R_p\right\} \quad \text{and} \quad \left\{(1-q_1) \le \mathrm{ucb}(1-\hat{q}_1)\right\}$$

each be events of probability $\sqrt{1-\alpha}$. Because of the independence, their joint occurrence has probability $1-\alpha$. It follows that

II. $$\Pr\left\{\mathrm{lcb}\left(\hat{R}_p\right) / \left[\mathrm{ucb}(1-\hat{q}_1)\right]^{k-2} \le R\right\} \ge 1-\alpha \tag{15}$$

and hence that the ratio in the above event, denoted $\mathrm{lcb}\left(\tilde{R}\right)$, is a lower confidence bound of probability at least $1-\alpha$. Gleser's numerical example shows that this bound can be a noticeable improvement over the earlier ones. See Table 2.

NPS.

This method was motivated by the fact that no item is subjected to more than one stress. So, the requirement of the device being able to function after being subjected to all of the stresses is not reflected in the testing program. The model adopted to use test results for this requirement is a highly conservative one. It says, in effect, that the device failure rate is the sum of the failure rates associated with the failure rates of items subjected to the single stresses. This form of accumulation is simple but overly conservative. Other models of damage accumulation can be rather complicated.

The author suggests that this requirement be abandoned in favor of one that is compatible with the test plan. Suppose that the maximum stress failure rate is used instead of the sum to represent the device failure rate. This can be estimated directly from the tests. It recognizes that the various stressful environments have their individual effects, but they are not additive. There is an extreme effect that dominates the others. Perhaps the

deterioration in system performance is represented by a worst case or cases; there may be many paths to those cases and the level of such deterioration would increase with the continued application of stresses. The test plan merely specifies an amount and variety of stresses.

This change in requirement is certainly far less conservative. It may be optimistic. It seems worthwhile to take a look at its effect.

I. The new reliability parameter is $R_0 = \max(R_2,\ldots, R_k)$. The empirical counterpart will be the point estimator

$$\hat{R}_0 = \max\left\{\hat{R}_2, \hat{R}_3, \ldots, \hat{R}_k\right\} \tag{16}$$

and an ad hoc lower confidence bound can be constructed as follows. Take the total sample size for the stressed tests

$$N_0 = \sum_{i=2}^{k} n_i \tag{17}$$

and the equivalent number of successes assuming the common success probability

$$X_0 = N_0 \hat{R}_0. \tag{18}$$

Treat $X_0 \sim \mathrm{Bin}(N_0, R_0)$ and use the incomplete Beta function method, Appendix B.

II. This method builds upon the log-linear analysis of the data (refined by failure category, see Section 6). The test plan, when viewed as a contingency table, shows that, on the basis of the data from tests that lead to product acceptance, there is no reason to claim that the treatment effects are not all the same. This supports a model that says the distributions in the columns of Table 3 are the same. In other words, we have no evidence that any of the $R_1, R_2,\ldots, R_k$ are different. The data can be pooled into a single set of 140 observations. Let RR be the probability of success and X be the total number of failures.

The point estimate of RR is $1 - X/140$ and the lower confidence bound follows from usual binomial theory.

5. Numerical Comparisons.

Let us apply all six methods to two data sets. They appear in Table 1. The first, DS1, was used in [6] and likely is fictitious. It has $k = 4$. The second, DS2, is a threshold case in the template of Appendix A; it also appears in Table 3. The three rows are sample sizes, failure counts, and empirical success rates.

Table 1. Data Sets

	Environment	1	2	3	4	5	6	7
DS1	Sample Size	20	20	32	20	–	–	–
	Failures	1	1	2	1	–	–	–
	$\hat{R}_i$	.95	.9375	.95	.95	–	–	–
DS2	Sample Size	50	5	5	20	20	20	20
	Failures	3	1	0	1	2	2	1
	$\hat{R}_i$	.94	.80	1.00	.95	.90	.90	.95

The application of the methods outlined produce the following results. (The DS1 results are not exactly the same as those in [6] because, using modern software, it is not necessary to interpolate from Incomplete Beta function tables.) The lower confidence level is 0.95.

Table 2

		Sandia		Gleser		NPS	
		I	II	I	II	I	II
DS1	$\hat{R}$	0.9461	0.9461	0.9375	0.9375	0.9375	0.9457
	lcb	0.3411	0.3681	0.3852	0.6186	0.8684	0.8891
DS2	$\hat{R}$	0.8848	0.8848	0.7969	0.7969	0.800	0.9286
	lcb	0.0686	0.1582	0.1753	0.1478	0.718	0.8819

Referring to DS1, the reliability estimates are generally high, but there is considerable dispersion in the lcb's. The Gleser techniques appear noticeably better than the Sandia ones. Although the NPS lcb's are even better, do not forget that they refer to less stringent

definitions of the device reliability. Turning to DS2, the reliability estimates are higher for the Sandia technique than for the Gleser. None of these two lcb's are high. DS2 may be special in that the observed maximal number of failures in the unstressed samples causes the reliability to appear higher than that for the stressed samples. It seems that this has an effect upon the lcb's as well. In contrast to this, the NPS values are attractive. But again one must not ignore the relaxation of the stringency of reliability definition.

Simple Modeling.

Although the test plan does not directly support the requirement that the reliability definition should include successful performance of the device after suffering all of the stresses, it may contain enough information to estimate model parameters that help describe that goal. The following describes a simple structure for accomplishing this. Suppose that we assert that

$$R_2 = R_1 d_2, \quad R_3 = R_1 d_3, \ldots, R_k = R_1 d_k \tag{19}$$

where R_1 is the probability that an unstressed device functions successfully and the $\{d_i\}$ represent the degradation of R_1 due to the i^{th} stress treatment. All d_i must satisfy the constraint that $0 \le d_i \le 1$ so a natural estimator would be

$$\hat{d}_i = \min\left(1, \hat{R}_i / \hat{R}_1\right) \tag{20}$$

and the estimator for the equation (1) definition of R would be

$$\hat{R} = \hat{R}_1 \prod_2^k \hat{d}_i . \tag{21}$$

The estimator (21) cannot be larger than (13). To see this, use (20) in (21) and rewrite it as

$$\hat{R} = \min\left(\hat{R}_1, \hat{R}_2\right) \min\left(\hat{R}_1, \hat{R}_3\right) \ldots \min\left(\hat{R}_1, \hat{R}_k\right) \Big/ \left(1 - \hat{q}_1\right)^{k-2} \tag{22}$$

and compare with (13). On the other hand, the variance of the numerator of (22) is smaller than that of (13). Heuristically, this should produce a larger lcb for the reliability using methods similar to Gleser I and II. Unfortunately the sampling distribution of this statistic is quite awkward and the lower confidence bounds difficult to find.

The situation improves slightly if we assume that all d_i are the same, say d. Then d could be estimated with an average of the d_i from equation (20) and the point estimator would become

$$\hat{R} = \hat{R}_1\left(\hat{d}\right)^{k-1} \tag{23}$$

but the lower confidence bound is still difficult.

Let us compute this point estimator for our two examples:

DS1: $\hat{R} = .9376$ DS2: $\hat{R} = .8$

In the second case the value is $\hat{R}_1$ because all $\hat{R}_i \geq \hat{R}_1$ for $i = 2,\ldots,7$. The point estimates are smaller but similar to those of the previous models. The advantage would appear in the lower confidence bound once such a computation has been developed.

6. Log Linear Analysis.

The result of testing each item in each group results in exactly one of the following four classifications mentioned in Section 2:

a&e. Color display or safety failure
b. Ignition failure
c&d. Delay or burning time specification failure
suc. No functioning out of zone; success.

The first three classifications are failure types and the fourth identifies a successful test. So far no use has been made of the extra information.

The test plan specifies the sample size for each group and threshold numbers for accepting the lot by classification for each group. Also there is a threshold failure count for the entire sampled set regardless of the group designation. For example, it may specify

20 signals for testing in the transportation & vibration group, and allow no a&e failures; at most one b failure; and at most two c&d failures for passage of the lot through that group. Requirements of this type are listed for all the groups and, in addition, there is a restriction on the total number of failures, by classification type, for the entire sampled set.

There are also a number of non destructive inspections, but they are of no concern in the development of reliability analysis for the sampling plan outlined.

Analysis.

It is recommended that the test plan be viewed as a two-way table of counts as exemplified in Table 3. (This is DS2 with greater detail.)

Table 3

	5ft	40ft	tran&vib	temp&hum	hi temp	lo temp	no stress	total
a&e	0	0	0	0	0	0	0	0
b	0	0	1	1	0	0	1	3
c&d	1	0	0	1	2	1	2	7
suc	4	5	19	18	18	19	47	130
total	5	5	20	20	20	20	50	140

Each column of this table is the result of multinomial sampling from a four-cell classification and the column total marks the sample size. The columns are statistically independent and are identified with the stress groups. Let $n(i, j)$ be the counts of items falling into the i^{th} classification in the j^{th} stress group, and let $p(i, j)$ be the multinomial probabilities for column j, $i = 1,\ldots,4$. The maximum likelihood estimator for the reliability of the j^{th} group is $n(4, j)/n(., j)$ where $n(., j)$ is the sample size for the j^{th} group. These are the same $\hat{R}_i$ numbers used earlier.

Let us fit a log-linear model, ref [1,4,5], to our two-way table. The general plan is to use the estimated cell probabilities from an acceptably fitted model. The minimum of the column reliabilities has been proposed for the estimated lot reliability. However the test

plan is such that the distinctions between the various stresses are not significantly different in the case of acceptable lots. (But this could change if the sample numbers were larger.) The common success probability is more readily estimated as done in NPS II. The log-linear model approach is described first.

Let $m(i, j)$ be the expected value of $\log\{n(i, j)\}$. The two-way layout model has been acceptably fitted for the test plans described above. In algebraic notation it is described by

$$m(i,j) = \mu + \lambda_i^A + \lambda_j^B \quad \text{for } i = 1\ldots4 \text{ and } j = 1\ldots7 \tag{24}$$

and linear contrast constraints [1,4] on the $\{\lambda_i^A\}$ and $\{\lambda_j^B\}$ apply.

The maximum likelihood estimates are easily found (either algebraically or with commonly available statistical software). Indeed the cell probabilities are the classical contingency table estimators. When the above log-linear model is tested against the saturated model, the fit cannot be rejected for data that are accepted by the test plans. Hence it follows that this procedure is the classical test for a common (multinomial) column distribution for the four classified outcomes. This in turn justifies the pooling of the column data into a single four-celled multinomial distribution and using the empirical probability of the fourth row as the estimated reliability. For the example above this value is

$$\hat{R} = \frac{130}{140} = .929.$$

The 95% lower confidence limit is .8819 (from usual binomial formulas). Finally, keep in mind that this example was chosen at the boundary threshold of acceptability according to the test plan rules.

The model (24) could have included some interaction terms of the form

$$\{\lambda_{ij}^{AB}\}.$$

A more detailed appropriate model could be developed. It would require more data in order to estimate its terms. Once done one can compute probabilities of the form

$$p_j(i) = \Pr\{\text{outcome } i \text{ given stress } j \text{ has been applied}\} = \exp\left\{\mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}\right\}. \quad (25)$$

A model of this type could be used to allocate the total sample over the various stress types and to specify the decision rule for each outcome category. One objective would be to avoid the condition of a small sample size for a stress type having undue influence on the estimates of system reliability. (Such has occurred for the five foot drop.)

It appears that new decision rules are being proposed; they do not allow any failures in order for a lot to be acceptable. It seems to this author that such a rule could produce considerable waste. The Operating Characteristic function is of the form

$$p^n \quad (26)$$

where p is the probability of a successful test and n is the number of items tested. Suppose p_1 is the lower limit specified for success. Values of p_1 around .95 and sample sizes of about 100 produce rather large Type I error probabilities. That is, many acceptable lots will be discarded.

It seems wiser to do more testing, keep better records, and develop sharper decision rules based upon careful modeling such as that suggested above. Then the process would be better understood and far fewer acceptable lots would be thrown out.

7. Summary.

The numerical comparisons in Table 2 for Sandia and Gleser serve to illustrate the conservative nature of the methods in place. The Gleser II method showed improvement for DS1, but seemed to break down when used with the DS2 data. The NPS approach shows attractive numbers but the interpretations are not the same.

The author argues that the test plan cannot support the original goal; at least not without some additional modeling. To begin with, no sample is subjected to all of the stresses and hence there is no direct testing of this condition. Further, the requirement does not specify the order of application of the various stresses. For example, suppose that

the vibration stresses are applied first. They hasten the loosening of the seals. This in turn makes the sample more susceptible to the effects of the temperature and humidity stresses. On the other hand, suppose that the vibration stresses are applied after the temperature and humidity stresses. Then the original resistance to temperature and humidity is maintained. Other scenarios reflecting the importance of order can be concocted.

The simple modeling in Section 5 may offer improvement, but it does nothing to offset the comments of the previous paragraph. The log linear analysis uses more of the test result information. If pursued, the results could be valuable for use in the design of test plans. The nature of the test plan can have substantial effects on the estimation of reliability. Recall that a single failure in the 5 ft drop stress can have a marked effect on the reliability estimate simply because of the low sample size.

Unfortunately a large amount of data must be analyzed in order to pick up any interaction information. Also there is likely to be a manufacturer effect that should be considered. Another idea would be to apply multiple stresses to the individual items. The log linear analysis could be quite useful if it were decided to do this. Perhaps simply applying them in pairs could reveal useful information.

At the present there appears to be no information relating the effect of sustaining a stress over a variety of time intervals, and not just those specified [12]. What is needed further is some information that can allow the modeling of the failure rates with time and how the various stresses compress time.

References

[1] Agresti, A (1990). *Categorical Data Analysis*, Wiley.

[2] Easterling, R.G. (1972). "Approximate Confidence Limits for System Reliability", *J. American Statistical Association*, **67** pp 220-222.

[3] Easterling, R.G., Mazumder, M., Spencer, F.W., & Diegert, K.D. (1991). *System-Based Component-Test Plans and Operating Characteristics: Binomial Data.* Technometrics.

[4] Fienberg, S.E. (1989). *The Analysis of Cross-Classified Categorical Data*, MIT Press.

[5] Freeman, G.H. & Halton, J.H. (1951). "Note on the Exact Treatment of Contingency, Goodness of Fit and Other Problems of Significance," *Biometrika*, **38**, pp 141-149.

[6] Gleser, L.J. Pyrotechnic Device Reliability Problem, Hand written communication.

[7] Gordon, W.O. & Taylor, D.M. (1986). *Reliability Methodology Implementation (QUATRA Model)*, Analysis & Simulation, Inc.

[8] Lloyd, D.K. and Lipow, M. (1962). *Reliability: management methods, mathematics.* Prentice-Hall.

[9] Maximus, Inc. (1980). Handbook for the Calculation of Lower Statistical Confidence Bounds of System Reliability.

[10] MIL-STD-1916 (1 April 1996). Department of Defense Test Method Standard.

[11] Myhre, J.M., Rosenfeld, A.M. & Saunders, S.C. (1978). "Determining Confidence Bounds for Highly Reliable Coherent Systems Based on a Paucity of Component Failures," *Naval Research Logistics Quarterly.*

[12] Naval Sea Systems Command, (1976). Prime Item Fabrication Specification. Signal, Smoke and Illumination, Marine MK 124 MOD 0.

[13] S-PLUS, Version 3.3, (1995). MathSoft, Inc.

[14] Spencer, F.W. & Easterling, R.G., *Lower Confidence Bounds on System Reliability Using Component Data: The Maximus Methodology. Reliability & Quality Control,* Elsivier Science Publishers B.V.

[15] Ushakov, I.A. (1994). *Handbook of Reliability Engineering*, Wiley.

APPENDIX A
Excerpts from Test Plan WS13697

The following table summarizes the decision rules for the destructive testing portion of the cited inspection plans, Table I (plan A). The notation $k/(k+1)$ means: 'accept with k or fewer failures/reject with $k+1$ or more failures'. The failure outcomes are as in Section 2.

	Stress Type	sample size	outcome	rule
i.	Five foot drop	5	a&e	0/1
			b,c&d	1/2
ii.	Forty foot drop	5	a&e	0/1
			b/c&d	1/2
iii.	Vibration	20	a&e	0/1
			b	1/2
			c&d	2/3
iv.	Temperature & Humidity	20	a&e	0/1
			b	1/2
			c&d	2/3
v.	High Temperature	20	a&e	0/1
			b	1/2
			c&d	2/3
vi.	Low Temperature	20	a&e	0/1
			b	1/2
			c&d	2/3
vii.	No Stress	50	a&e	0/1
			b	1/2
			c&d	3/4
	total	140		
	Group requirement (for the above tests)	140	b	3/4
			c&d	7/8

The group requirement places an additional limitation on the total acceptable number of failures, by failure outcome. (Of course no a&e failures are tolerated.)

APPENDIX B

The Use of the Incomplete Beta Function to Produce Binomial Confidence Limits

Suppose X is a binomial (n, p) random variable. The cumulative distribution function is related to the incomplete beta function by the well known formula

$$P(x) = P\{X \le x\} = \sum_{j=0}^{x} \binom{n}{j} p^j (1-p)^{n-j}$$

$$= 1 - \mathrm{IB}(p, x+1, n-x)$$

and the incomplete beta function is given by

$$\mathrm{IB}(p, x+1, n-x) = n\binom{n-1}{x} \int_0^p t^x (1-t)^{n-x-1} dt.$$

The Maximus [9] methodology employs a technique of identifying subsystems of components with a fictional equivalent component. The reliability of the subsystem is estimated from the results of testing its components and the result is viewed as a binomial proportion. But the parameters of this equivalent component need not be integral numbers. That is, neither the sample size nor the number of successes need be whole numbers. This leads to the need for some interpolation formulae. Incomplete Beta function software is readily available, so one may as well use it. The coefficient in IB is related to the Beta function

$$n\binom{n-1}{x} = [\mathrm{Beta}(x, n-x-1)]^{-1}$$

and there is no longer any requirement that x and n be integers.

To find a $1-\alpha$ level lower confidence limit for p one seeks the solution of

$$\alpha = \sum_{j=x}^{n} \binom{n}{j} p^j (1-p)^{n-j} = \mathrm{IB}(p, x, n+1-x)$$

as a function of p. When $x = 0$, use zero for the lcb.

To find a $1 - \alpha$ level upper confidence limit for p one seeks the solution of

$$\alpha = \sum_{j=0}^{x} \binom{n}{j} p^j (1-p)^{n-j} = 1 - \mathrm{IB}(p, x+1, n-x).$$

When $x = n$, use unity for the ucb. Software for computing the inverse incomplete beta function is available. For example, the function qbeta($p, x + 1,\ n - x$) in the SPLUS package [13] is used in our work.

APPENDIX C
The Variance of a Product of Independent Random Variables

Let $\mu_i = E(X_i)$ and $\sigma_i^2 = V(X_i)$, for $i = 1, 2, \ldots$. The formula for the variance of the product is:

$$Var\left(\prod_{j=1}^{k} X_j\right) = E\left\{\prod_{1}^{k} X_j^2\right\} - \prod_{1}^{k}\left(\mu_j^2\right)$$

$$= \prod_{1}^{k}\left(\sigma_j^2 + \mu_j^2\right) - \mu_1^2\mu_2^2 \ldots \mu_k^2.$$

From this one can verbalize the expanded series in the following way. It is a series of terms having k factors per term. Each factor is of second order, i.e. either a μ^2 or a σ^2 (with subscripts). Every combination of μ^2 and σ^2 must appear; there is a term consisting of the product of all variances; there is no term which is the product of all squared means, so $2^k - 1$ terms in total.

For our application the X_i are binomial proportions so

$$\mu_i = p_i \quad \text{and} \quad \sigma_i^2 = p_i(1-p_i)/n_i \quad \text{for } i = 1, \ldots k.$$

If the $\{n_i\}$ are large then we can ignore all terms that have two or more of the $\left\{\sigma_i^2\right\}$. This leads to the simplified approximation

$$V(\hat{p}_1\hat{p}_2 \cdots \hat{p}_k) = \sum_{i=1}^{k} \frac{p_i(1-p_i)}{n_i} \prod_{j \neq i} p_j^2$$

$$= \left(\prod_{1}^{k} p_i\right) \sum_{i=1}^{k} \frac{p_1 p_2 \cdots p_{i-1} p_{i+1} \cdots p_k (1-p_i)}{n_i} \tag{C.1}$$

All of the ignored terms are positive.

APPENDIX D
Details for a Delta Method

We develop the delta method for approximating the variance of

$$W = X/Y^r \tag{D.1}$$

assuming X and Y are independent random variables and r is an integral constant.

Start with the first order Taylor expansion of $W(X, Y)$ about (x_0, y_0).

$$W = x_0 / y_0^r + (X - x_0)/y_0^r - rx_0(Y - y_0)/y_0^{r+1}.$$

The remainder terms are assumed negligible when (x_0, y_0) is the centroid of the (X, Y) distribution. Then

$$Var(W) \approx Var(X)/y_0^{2r} + \left(x_0^2 r^2 / y_0^{2r+2}\right) Var(Y). \tag{D.2}$$

The present application calls for $X = \hat{R}_p$, $Y = (1 - \hat{q}_i)$ and $r = k - 2$. Then

$$Var\left\{\hat{R}_p / (1 - \hat{q}_1)^{k-2}\right\} \approx \frac{Var\left(\hat{R}_p\right)}{(1 - \hat{q}_1)^{2k-4}} + (k-2)^2 \hat{R}_p^2 \frac{Var(1 - \hat{q}_1)}{(1 - \hat{q}_1)^{2k-2}} \tag{D.3}$$

APPENDIX E
Alteration of Empirical Proportion Estimates

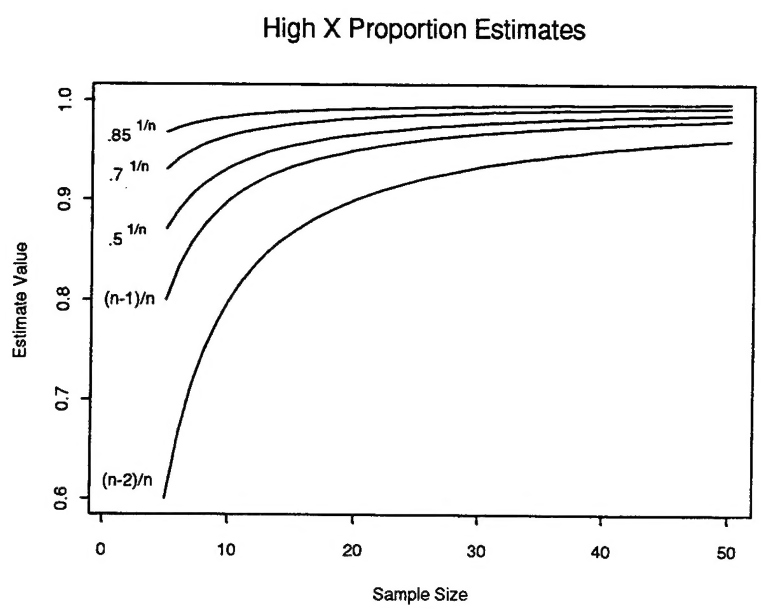
When X is distributed $\sim$ Bin(n, p) the maximum likelihood estimator for p is the empirical proportion of successes. This produces values of zero or one in the extreme cases of $X = 0$ or $X = n$, respectively. In many applications such values are either inappropriate or unusable and relief from these extremes is required. A popular form of relief is to use one sided confidence limits as point estimates in these two cases. In the case $X = n$ the $1 - \alpha$ lower confidence limit for p is computed from the equation

$$p^n = \alpha .$$

and in the case $X = 0$ the $1 - \alpha$ upper confidence limit is computed from

$$(1 - p)^n = \alpha.$$

The Maximus technology [9] uses $\alpha = 0.7$. The following graphs provide some visual comparisons of estimators of this type by plotting the estimators for several choices of α as a function of n. Also included are some maximum likelihood estimate curves for values of X 'nearby'. The '$X = 0$' curves are simply one minus the '$X = n$' curves.



INITIAL DISTRIBUTION LIST

1. Research Office (Code 09) 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

2. Dudley Knox Library (Code 013) 2
   Naval Postgraduate School
   Monterey, CA 93943-5002

3. Defense Technical Information Center 2
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, VA 22060-6218

4. Therese Bilodeau 1
   Dept of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

5. Prof. Robert R. Read (Code OR/Re) 7
   Naval Postgraduate School
   Monterey, CA 93943-5000

6. Prof. Max Woods (Code MA/Wm) 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

7. Commander, ... 1
   NAVSURFWARCENDIV Crane
   (Code 0563, Bldg 2701)
   300 Highway 361
   Crane, IN 47522-5001